Skill Overshooting in Job Training with the Trade Adjustment Assistance Program

Online Appendix

Justin Barnette
Department of Economics
Kent State University
Email: jbarne25@kent.edu
Tel. 330 - 672 - 1096

Jooyoun Park
Department of Economics
Kent State University
Email: jpark8@kent.edu
Tel. 330 - 672 - 1086

July 13, 2020

Abstract

We investigate training choices made by workers entering the Trade Adjustment Assistance (TAA) program. This is important as more workers enter these types of programs due to technological change and globalization. We show that workers that choose a training occupation beyond their skill level (skill overshooting) achieve higher earnings and wage replacement rates with the cost being that it lowers their reemployment rates. Specifically, skill overshooting lowers the reemployment rates for these trainees by 2.0 percentage points, but they enjoy an increase in their wage replacement rate by 2.0 percentage points and $615 in annual earnings. An investigation of subsamples, shows that skill overshooting affects different groups of trainees differently. Female and rural-dwelling trainees enjoy the most benefits in earnings ($1,443 and $1,080, respectively) without hurting their chances of reemployment. The highly educated sample enjoys a large increase in earnings but still bears the decline in the reemployment rate.

Keywords: TAA, job training, wage replacement, ALMP
JEL: J08, J68, F16

A1. Preparing earnings record for estimation

TAPR reports participants earnings prior to participation and post program exit. Pre-participation earnings are recorded for three quarters prior to participation. These three quarters are calendar quarters prior to participation rather than prior to separation. This makes some of the earnings records invalid, especially in terms of constructing the wage replacement rate.

We use $2,000 for the lower bound and $50,000 for the upper bound for earnings record validity. The lower bound is chosen to mimic the minimum earnings requirement for Unemployment Compensation eligibility. The minimum requirement values differ across states and change over time, but the figure is comparable to the weekly earning of a minimum wage worker working for 30 hours. All our earnings records are converted to 2000 USD value using the national consumer price index. In 2000, the federal minimum wage was $5.15. 30 hours per week for 13 weeks at that rate makes the quarterly earnings $2,008.5. We choose $2,000 for the lower-bound. The upper bound of $50,000 is chosen somewhat artificially based on the data. As one can expect, TAA participants – mostly comprised of manufacturing workers – do not often fall into the high-income category. A large number of participants have reported quarterly earnings of $25,000 or below. $50,000 is chosen to limit the influence of outliers without affecting the distribution of earnings too much.

Table A1 shows the statistics regarding the reported values of quarterly earnings before and after participation. It shows that a non-negligible fraction of participants reported quarterly earnings below $2,000 for all six quarters used for earnings reporting, but the upper bound of $50,000 does not account for a meaningful share of our sample. Comparing panels, A and B of Table A1, shows that the standard deviation of quarterly earnings declined drastically after imposing the upper and lower bounds. This is probably largely from dropping reported values beyond the upper bound rather than the lower bound.

Then we construct the pre-participation earning by taking the maximum value among the valid earnings record out of three quarterly earnings. Lastly, we check when participants were separated from the qualifying position. We drop the earnings record if the qualifying separation occurred 5 quarters or

Table A1. Cleaning of Reported Quarterly Earnings

	Obs	Mean	Std. Dev.	Min	Max	# below $2,000	# above $50,000
A. RAW DATA							
Pre-participation							
Three quarters prior	281,962	10,067.8	47,902.0	0.01	1,000,000	12,823	1,146
Two quarters prior	279,274	10,210.2	49,629.8	0.80	1,000,000	16,365	1,345
One quarter prior	121,448	15,552.5	81,682.4	0.06	1,000,000	14,863	1,631
Post-exit							
First quarter post	216,694	5,072.9	3,895.0	0.1	119,663	39,096	63
Second quarter	222,167	5,375.5	4,333.2	0.0	805,744	32,347	69
Third quarter	222,312	5,565.5	3,980.1	0.0	231,713	29,560	58
B. CLEANED DATA							
Pre-participation							
Three quarters prior	267,993	7,675	4,372	2,000	49,965		
Two quarters prior	261,564	7,691	4,612	2,000	49,983		
One quarter prior	104,954	8,559	6,006	2,001	49,987		
Pre-participation earnings	284,368	8,739	5,599	2,000	49,987		
Post-exit							
First quarter post	177,535	5,950	3,568	2,000	49,919		
Second quarter	189,751	6,092	3,539	2,000	49,471		
Third quarter	192,694	6,245	3,626	2,000	49,902		
Post-participation earnings	220,141	6,700	4,060	2,000	49,919		

more before participation. Three quarterly earnings following the program exit are treated in the same manner with lower bound of $2,000 and upper bound of $50,000.

A2. Occupation Codes Reporting Quality

Occupation codes in TAPR are reported using the 8-digit Standard Occupational Classification (SOC) system. The SOC system is structured with the first two digits representing occupation families. The following four digits describe specific occupations and the last two digits describe subcategories of an occupation.[36] We use the 6-digit SOC as our occupation classification but these codes are not always provided. Among 320,603 participants observed in our sample, 209,762 received occupational training

[36] Many 8-digit SOC codes end with 00.

(classroom, customized, or on-the-job training) and of those 173,012 have a training occupation reported with 118,925 having a valid[37] 6-digit code. Reemployment occupation reporting is reported less. 247,511 participants are indicated as employed at least for one quarter during the three-quarter observation period following the exit of the TAA program. 87,410 reported an occupation of which 66,113 have a valid 6-digit SOC. Overall, 56.70 % of participants who received occupational training have a valid occupation code. This share is 26.71 % for reemployment occupations.

The reporting quality does not seem to be participant specific as some participants have one reported and valid and the other missing. A total of 168,793 participants received job training and found a job after exiting the program. Only 43,064 (25.51%) have valid occupation codes for both training and reemployment occupations. 53,507 have only the training occupation, 9,874 have only the occupation of reemployment, and 62,348 have neither.

A3. IPUMS CPS Education Variable Construction

In the construction of skill overshooting and skill upgrading variables, we use the average education attainment for each occupation from IPUMS CPS to be compared to the participant's own level of education. We construct the occupation-level average education by taking the average years of schooling for job holders of each occupation in using IPUMS CPS from 1998 to 2008. Skill overshooting uses the occupation-level education for the year of participation and skill upgrading uses that for the year of program exit. IPUMS CPS reports the occupation of each observation using Census Occupation Codes (COCs). We first crosswalk COCs to SOC codes using 2000 crosswalk file published by BLS then merge the occupation level education variable to TAPR using SOC codes.

Table A2 presents the most frequently observed occupations for training and reemployment and Table A3 shows statistics for schooling for relevant individuals from both TAPR and CPS for five selected occupations that are most frequently observed in TAPR. This shows how CPS individuals

[37] By valid codes we mean that the reported occupation code has at least 6 digits and the first two digits are odd numbers between 11 and 55.

compare to TAA participants. For all five occupations, individuals with various educational background are working for the same occupation. For instance, workers with 7 years of education to those with and advanced degree are holding a job as a truck driver. The mean values of schooling also do not differ too much across the TAPR samples (those who choose the occupation for training and those who find a job in that occupation) and the CPS sample. The CPS sample shows a slightly higher average years of schooling for some occupations (e.g. general office clerks and medical assistants).

In the analysis, education values from TAPR are used individually. This preserves variations from 7 to 17 years in educational attainment for the same occupation. The value from CPS used in the analysis is only the mean value for each occupation which varies from 10.5 to 17.0.

Table A2. Most popular occupations of training and reemployment

I. TRAINING	
SOC	**SOC title**
533032	Truck Drivers, Heavy and Tractor-Trailer
319092	Medical Assistants
311012	Nursing Aides, Orderlies, and Attendants
439061	Office Clerks, General
499021	Heating, Air Conditioning, and Refrigeration Mechanics and Installers
439011	Computer Operators
292061	Licensed Practical and Licensed Vocational Nurses
151041	Computer Support Specialists
436013	Medical Secretaries
292071	Medical Records and Health Information Technicians
II. REEMPLOYMENT	
SOC	**SOC title**
533032	Truck Drivers, Heavy and Tractor-Trailer
311012	Nursing Aides, Orderlies, and Attendants
519199	Production Workers, All Other
519198	Helpers--Production Workers
319092	Medical Assistants
499021	Heating, Air Conditioning, and Refrigeration Mechanics and Installers
439061	Office Clerks, General
519061	Inspectors, Testers, Sorters, Samplers, and Weighers
516031	Sewing Machine Operators
537062	Laborers and Freight, Stock, and Material Movers, Hand

** Minimum and Maximum education for these occupations are mostly 7 years and 17 years with two occupations with minimum education of 8 years and two occupations with maximum education of 16 years.

Table A3. Average educational attainment of popular occupations

		Training Occ (OSTC)			Reemployment Occ (OCE)			CPS		
SOC	Occ title	Obs.	mean	min/ max	Obs.	mean	min/ max	Obs	Mean	min/ max
533032	Truck Drivers, Heavy and Tractor-Trailer	6,963 6.79%	12.00 (1.33)	7/17	2,646 4.41%	12.06 (1.37)	7/17	259,633	12.17 (1.47)	7/17
311012	Nursing Aides, Orderlies, & Attendants	4,164 4.06%	11.61 (1.43)	7/17	2,074 3.46%	11.84 (1.27)	7/16	162,342	12.50 (1.53)	7/17
319092	Medical Assistants	4,711 4.60%	12.19 (1.19)	7/17	2,469 2.45%	12.42 (1.15)	8/17	51,039	13.03 (1.44)	7/17
439061	Office Clerks, General	4,070 3.97%	11.66 (1.65)	7/17	1,081 1.80%	12.51 (1.24)	8/17	102,721	13.13 (1.57)	7/17
499021	Heating, Air Conditioning, and Refrigeration Mechanics and Installers	3,930 3.83%	12.37 (1.26)	7/17	1,126 1.88%	12.51 (1.18)	7/17	26,519	12.50 (1.35)	7/17
	All	123,852	12.40 (1.51)	7/17	72,269	12.49 (1.56)	7/17	13,339,017	13.45 (2.10)	7/17

A4. Identification of the Urban-Rural Nature of Commuting Zones

For the analysis on the geographic subsample analyses, we divide our sample to trainees living in urban areas and rural areas with the unit of local area being commuting zones (CZs). We identified urban/rural nature of each CZ based on urban-rural continuum code (2003 version) published by U.S. Department of Agriculture. This is reported at the county-level and we aggregate these county-level codes to the CZ-level.

The continuum code places each county on a 9-point urban-rural spectrum with 1 as the most urban and 9 the most rural based on three factors: first, whether a county is a metro area; second, population of the county; third, whether a non-metro county is adjacent to a metro county. A metro county is divided into three codes based on the population (code 1 = 1 million population or more, code 2 = 250,000 to 1 million population, code 3 = 250,000 or fewer than 250,000). Non-metro counties are divided by urban population thresholds of 20,000 and 2,500. Then each population category is divided into two based on the adjacency to a metro area. A county with the continuum code 9 is described as either completely rural or less than 2,500 urban population and not adjacent to a metro area.

A CZ is classified as urban if all counties in the CZ are either a metro area or adjacent to a metro area. Out of 708 CZs in the U.S., 298 are classified as urban based on this criteria. The average population of urban CZs is 783,362 based on 2000 Census. Among 320,603 participants in our TAPR dataset, 191,081 are identified to reside in an urban CZ. Defining an urban commuting zone as a commuting zone with all counties in the metro area or non-metro counties that are adjacent to a metro area means that no county in the CZ is too isolated and all counties are within a reasonable distance from a metro area. The urban-ness of a commuting zone is broadly defined and an urban CZ should not be interpreted as a densely populated metro area.

A CZ is classified as rural if no county in the CZ is identified as a metro area. 397 out of 708 CZs are classified as rural. The average population of these CZs in 2000 Census is 63,659 which is less than 10% of the size of an average urban CZ. 49,486 observations in our TAPR sample are identified to reside in a rural CZ compared to 191,081 TAPR observations in urban area. i.e. The vast majority of TAA participants reside in an urban area.

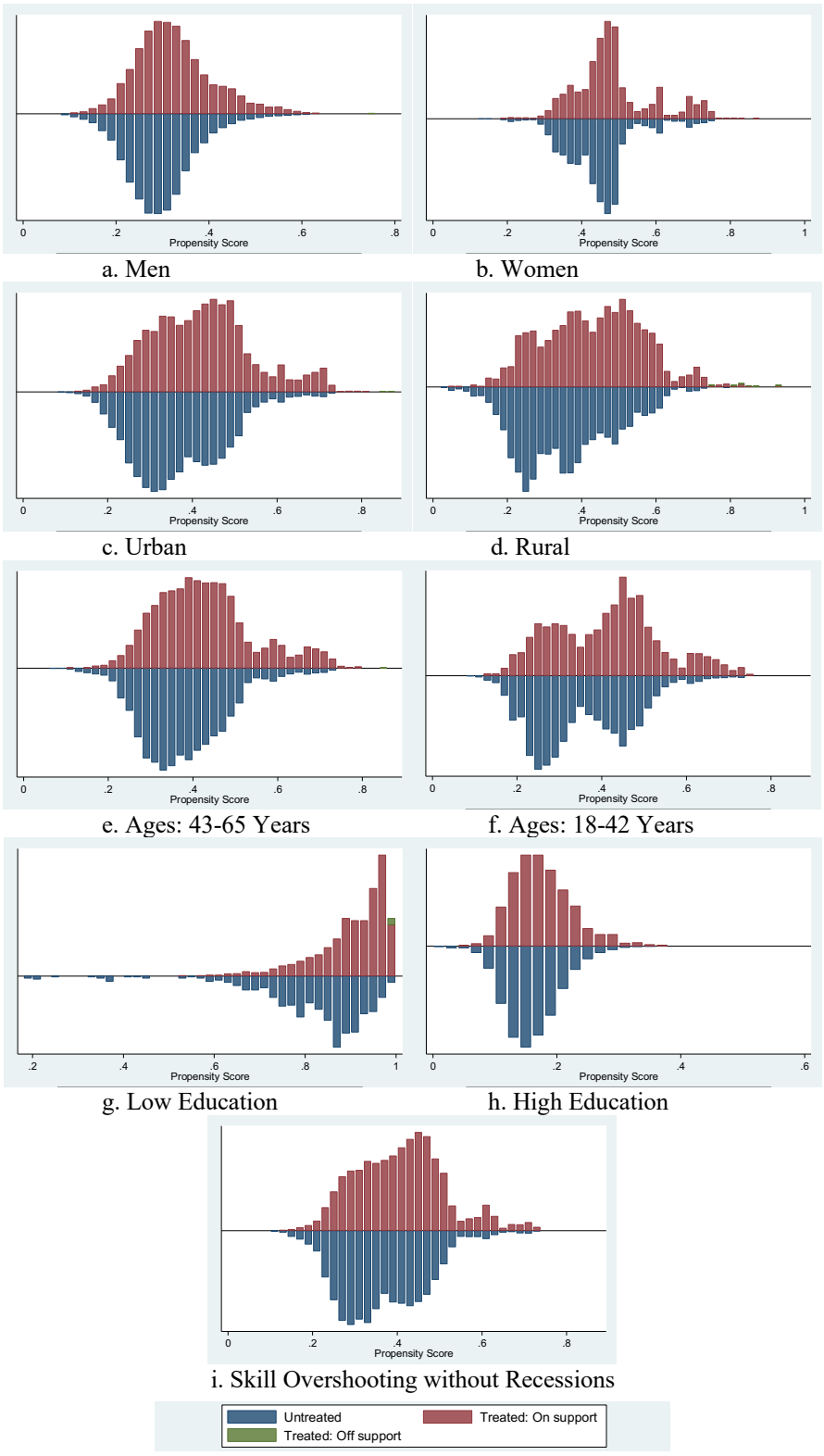
A5. Balancing Test

Table A4 demonstrates the bias before and after the first stage of matching. The t-tests indicate that all but one variable was statistically different before matching but matching reduces this bias to statistically insignificant differences. Before matching, the comparison group had it strongest bias in gender. Initially, the data on those that do not overshoot was made up of 55.28% male trainees, compared to the treated group or the overshooters who were 38.51% male. After matching and the enforcement of the common support as discussed above, the treated group is only slightly adjusted but the comparison group is now 38.66% male. Given the large differences in gender that we discussed in the main body of the paper, this 99.2% reduction is crucial. Notice similar reductions throughout the rest of the variables as indicated in the table.

Table A4. Treated vs. Comparison Group Matching Quality Statistics

		Mean			Reduction	t-test	
Variable	Matched	Treated	Comparison	% Bias	in Bias (%)	t	p>|t|
Gender: Male	Before	.38507	.55275	-34.1		-38.21	0.000
	After	.3853	.38666	-0.3	99.2	-0.28	0.777
Pre-Part Earnings	Before	31642	36077	-21.4		-23.70	0.000
	After	31648	31750	-0.5	97.7	-0.54	0.587
Limited English	Before	.06294	.03599	12.5		14.42	0.000
Proficiency	After	.06254	.06626	-1.7	86.2	-1.54	0.124
Tenure	Before	9.8573	10.349	-5.6		-6.22	0.000
	After	9.8586	9.8469	0.1	97.6	0.14	0.891
Age at	Before	42.639	43.274	-6.4		-7.16	0.000
Participation	After	42.643	42.625	0.2	97.1	0.19	0.852
Eth: Black	Before	.15143	.16824	-4.6		-5.13	0.000
	After	.15151	.15159	-0.0	99.6	-0.02	0.984
Eth: Hispanic	Before	.15215	.09164	18.6		21.41	0.000
	After	.15166	.15304	-0.4	97.7	-0.39	0.696
Eth: Asian	Before	.02399	.03257	-5.2		-5.72	0.000
	After	.02401	.02595	-1.2	77.3	-1.26	0.206
Eth: Other	Before	.01166	.01191	-0.2		-0.26	0.795
	After	.01166	.0112	0.4	-84.4	0.44	0.660

Figure A1. Predicted Probabilities of Skill Overshooting for Subsamples



Note: These come from the first part of matching in the estimation for the subsamples of Section V(c). *Treated: On support* provides our sample of those that overshoot which are matched from two of the comparison group labeled *Untreated*. The label, *Treated: Off support*, are those that overshoot but have no match. The subsamples in figures a and b are split by gender. The subsamples in figures c and d are split by geographic location using the Urban-Rural Continuum Code discussed in section III(d). The subsample in g has less than 12 years of education and the subsample in h has more than 12 years of education. The subsample in figure g are for those that exited training in 2003-2007.